

## Analyse Sozialer Medien an der Schnittstelle zwischen Informatik und Sozialwissenschaften

Strohmaier, Markus; Zens, Maria

Veröffentlichungsversion / Published Version  
Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Strohmaier, M., & Zens, M. (2014). Analyse Sozialer Medien an der Schnittstelle zwischen Informatik und Sozialwissenschaften. In C. König, M. Stahl, & E. Wiegand (Hrsg.), *Soziale Medien : Gegenstand und Instrument der Forschung* (S. 73-95). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-05327-7\\_4](https://doi.org/10.1007/978-3-658-05327-7_4)

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der  
  
Leibniz-Gemeinschaft

# Analyse Sozialer Medien an der Schnittstelle zwischen Informatik und Sozialwissenschaften

Markus Strohmaier und Maria Zens

GESIS – Leibniz-Institut für Sozialwissenschaften

Ihre Bezeichnung zeigt es bereits an: Die Sozialen Medien – Twitter, Facebook, YouTube, Google+ und eine Vielzahl weniger bekannter Angebote, Community-Portale oder Blogs – sind unbestritten ein Interaktionsraum mit immenser sozialer Bedeutung. Eine (beinahe) ubiquitäre Zugänglichkeit, die große Reichweite und ein hoher Durchdringungsgrad würden sie allein in quantitativer Hinsicht zu einem beachtenswerten Phänomen machen.<sup>1</sup> Sie tragen aber auch zu einem „neuen Strukturwandel der Öffentlichkeit“ bei – wie die drei deutschsprachigen Fachgesellschaften für Soziologie aus Deutschland, Österreich und der Schweiz Ende 2011 ihren Dreiländerkongress in Anknüpfung an die knapp fünfzig Jahre zuvor erschienene bahnbrechende Studie von Jürgen Habermas betitelten.<sup>2</sup>

Auch die wissenschaftliche Wahrnehmung der Sozialen Medien hat sich verschoben – von einer Freizeitbeschäftigung Jüngerer zum universellen Kommunikations- und Speicherraum, der individuellen und kollektiven Akteuren, Privatleuten, Firmen und Organisationen gleichermaßen offen steht. Die Besonderheit der Sozialen Medien ist, dass sie auf den Beiträgen der Benutzer (*user generated content*) basieren und damit auch Informationen zu allen möglichen Lebensbereichen enthalten, die kaum professionell oder institutionell gefiltert würden oder den Anspruch hätten, Teil einer soziopolitischen Öffentlichkeit zu sein.

Aus sozialwissenschaftlicher Sicht bedeutet dies eine Entgrenzung von Privatem und Öffentlichem, die eigene Kommunikationsformen generiert

---

1 Cf. Alexa; <http://www.alexa.com/topsites>

2 Cf. zum Programm <http://www.soziologie2011.eu/>

und einen neuen Informationshabitus prägt. Dieser selbst ist ein Untersuchungsgegenstand sozialwissenschaftlicher Forschung – von der Vernetzung über digitale Ungleichheiten bis zur Konstitution politischer Arenen in den neuen Medien. Die Geschwindigkeit des kommunikativen Wandels wird deutlich, sobald man sich vor Augen hält, dass Google gerade einmal 15 Jahre besteht und vor zehn Jahren weder Facebook noch Twitter oder YouTube existierten. Die Effekte des Wandels sind eine Herausforderung: Als Quelle für sozialwissenschaftlich relevante Daten bieten die Sozialen Medien zwar eine Überfülle von Informationen (*Big Data*), aber zunächst einmal keine hoch-qualitativen Daten im herkömmlichen Sinn.

Das Ziel dieses Beitrags ist es, gemeinsame Interessen, Möglichkeiten und Notwendigkeiten für Kollaborationen von Computer- und Sozialwissenschaften in diesem neu entstehenden Feld aufzuzeigen, und inwiefern die einen vom Fachwissen der jeweils anderen profitieren können. Adressaten sind deshalb vor allem Wissenschaftler an der Schnittstelle von Informatik und Sozialwissenschaften, aber auch die breitere Fachöffentlichkeit, die sich für die gesellschaftliche Relevanz von *Big Data* und die damit einhergehende Herausforderung an die Erhebungs- und Auswertungsverfahren interessiert.

Im ersten Abschnitt werden wir in einem kurzen Rekurs auf aktuelle Debattenbeiträge zu dieser Frage eingehen und skizzieren, wo die Herausforderungen für die Sozialwissenschaften liegen. Anschließend sollen anhand einer Reihe von Beispielbereichen mögliche, sozialwissenschaftlich relevante Analysen von Transaktionsdaten aufgezeigt werden. Damit soll deutlich werden, welche Informationen in den Sozialen Medien gespeichert sind und wie sie für die soziologische, politikwissenschaftliche, kommunikationsanalytische usw. Forschung nutzbar gemacht werden können. Hierzu gehören die Netzwerkforschung (Abschnitt 2) oder die Nutzung von Sozialen Medien für die politische Aktivierung und Information (3), aber auch die genauere Beschäftigung mit der intentionalen Struktur von Suchanfragen (4). Kritische Betrachtung verdienen die Effekte von Trendverstärkung und Selektion durch Suchmaschinen bis hin zu sogenannten *Social Bots*, wie sie in den Abschnitten 5 und 6 angesprochen werden. Zum Abschluss (7) soll die „Konfluenz“ von computer- und sozialwissenschaftlichen Perspektiven auf diese Fragestellungen aufgezeigt werden.

## 1 Big Data + Internet Time = Krise der Sozialwissenschaften?

Die beiden Soziologen Mike Savage und Roger Burrows annoncierten 2007 eine aufziehende Krise der Sozialwissenschaften. Krisen sind der Disziplin vertraut; hier ging es jedoch nicht um eine durch Selbstreflexion und methodologische oder gesellschaftliche Verwerfungen induzierte Positionierungs- und Legitimationskrise, sondern um ein strukturelles Problem, das durch die Proliferation und schiere Menge relevanter Daten im Web erzeugt werde und den – für lange Zeit als stabil eingeschätzten – Kern der empirischen Sozialforschung und sozialwissenschaftlichen Datenerhebung infrage stelle (Savage/Burrows 2007).

Die klassischen Verfahren der empirischen Soziologie – repräsentative Umfragen, Interviews – so die Argumentation, seien kaum in der Lage, das Soziale, wie es sich in diesen neuen Formen artikuliert, zu erfassen. Programmatisch heißt es „both the sample survey and the in-depth interview are increasingly dated research methods, which are unlikely to provide a robust base for the jurisdiction of empirical sociologists in coming decades“ (Savage/Burrows 2007, 885). Für die beiden Soziologen sind damit die „glorreichen Tage“ der Umfrageforschung vorbei, deren Stelle könnten digitale Daten, wie sie als Nebenprodukte ihrer Transaktion anfallen, einnehmen. Es ist kaum zu leugnen: Diese Daten sind einfacher, kostengünstiger und schneller zu erhalten als in Langzeitstudien oder Tiefeninterviews erworben; sie bergen aber auch Nutzungsrisiken und haben qualitative Grenzen, die derzeit kaum erforscht sind. Hierzu gehören die Fragen nach Repräsentativität, den Standards der Erhebung, Bereinigung und Interpretation, aber auch ganz neue wie z.B. die im Verlauf der Sammlung und Evaluation maschinen- von menschengenerierter Kommunikation unterscheiden zu müssen.

Die provokante Publikation von Savage und Burrows hat eine ganze Reihe von unmittelbaren Reaktionen hervorgerufen<sup>3</sup> und darüber hinaus eine breite Diskussion in Gang gesetzt, die zunehmend die methodologischen Herausforderungen von *Big Data* – Datenmenge – und *Internet Time* – Veränderungsgeschwindigkeit – adressiert.

---

3 Eine Sammlung findet sich 2012 als erstes E-Issue von *Sociology* unter dem Titel „Exploring Trends and Challenges in Sociological Research“, herausgegeben von Linda McKie und Louise Ryan.

Tom Smith zeichnet in einem aktuellen Beitrag zu „Survey-Research Paradigms Old and New“ Grundlinien der Diskussion um diese doppelte Herausforderung von Quantität und (mangelnder) Qualität nach: „They [social media] provide information about public opinion and are a source of information about many other societal aspects, but are generally more limited than standard surveys in capturing the full range and complexity of public opinion and measuring society’s values and behaviors“ (Smith 2013, 8).

Mit derselben Konstellation befasst sich David Karpf in „Social Science Research Methods in Internet Time“ (2012), kommt aber zu etwas anderen Einschätzungen; er betrachtet die Internetforschung als „messy-but-promising“, als genuin trans- und interdisziplinäres Unternehmen. Für vielversprechend hält Karpf insbesondere den Trend hin zu einer *Computational Social Science*, die sich allerdings der qualitativen Grenzen ihrer Daten stets bewusst sein müsse. Statt einer Rückkehr zu den überkommenen robusten Methoden der Sozialforschung plädiert er hinsichtlich der Erforschung von Web-Daten für eine „partial solution: embracing transparency and kludginess“, für die Kombination von *Computational Social Science* und qualitativen Fallstudien, die dazu beitragen, die erhobenen Datenmengen auf die zugehörigen sozialen Phänomene zu beziehen, in ihren Bedeutungen erklären und prüfbare Hypothesen bilden zu können (Karpf 2012, Zitate: 18,14).

Ob man es nun als Dilemma oder Chance begreift, als Ersatz, Re-Orientierung oder Verbesserung (Edwards et al. 2013 zu diesen Perspektiven), gerade die Forderung nach Datenqualität zeigt: Die Sozialwissenschaften benötigen informatische Verfahren und computergestützte Prozesse, um Internet-Transaktionsdaten nach ihren eigenen Standards extrahieren und beforschen zu können.

Im Folgenden sollen mit einigen Beispielen Möglichkeiten der „Konfluenz“ informatischer und sozialwissenschaftlicher Perspektiven aufgezeigt werden, soll auf bereits vorhandene Schnittmengen und Anknüpfungspunkte hingewiesen werden. Hierzu gehören beispielsweise Konzepte von Vernetzung, bei denen soziologische Ansätze und mathematisch-statistische Verfahren ineinandergreifen, oder die Analyse soziopolitisch relevanter Inhalte und Interaktionen im Web. In der Regel wird es so sein, dass die Sozialwissenschaften für sie relevante Fragestellungen auf Web-Inhalte und -Interaktionen anwenden müssen, während die Informatik dafür zuständig ist, diese Informationen bzw. Interaktionsspuren zu extrahieren.

## 2 Modellbildung: Strukturen der Vernetzung

Eines der wohl populärsten Konzepte der sozialen Vernetzung ist das *small world phenomenon*. Es geht auf Arbeiten des US-amerikanischen Sozialpsychologen Stanley Milgram vom Ende der 1960er Jahre zurück. Milgram geht davon aus, dass wir alle über relativ kurze Wege verbunden sind; oder, anders formuliert: Jeder ist von jedem nur durch eine erstaunlich geringe Zahl an persönlichen Kontakten getrennt (in der Folge, allerdings nicht von Milgram selbst, oft auch als *six degrees of separation* bezeichnet). Milgram überprüfte seine Hypothese von der „kleinen Welt“, indem er zufällig ausgewählten Personen aus Omaha und Wichita die Aufgabe stellte, ein Päckchen an in Boston lebende Zielpersonen zu senden – sofern diese ihnen persönlich unbekannt war, sollte dies nicht auf direktem Weg geschehen, sondern eine andere, ihnen bekannte Person ausgewählt werden, von der sie vermuteten, sie könnte die Zielperson kennen etc. Die als „Zwischenstation“ fungierenden Personen waren jeweils aufgefordert, zur Pfad-Dokumentation Postkarten an die Forscher zu senden. Im Ergebnis benötigten die erfolgreichen Zustellungen rund 6 Stationen (Milgram 1967; Travers/Milgram 1969).

Diese von Milgram vor bald einem halben Jahrhundert aufgestellte These von der „kleinen Welt“ wurde in der Folge durchaus kontrovers diskutiert; kritisiert wurden u.a. die Anlage des Experiments, die zu geringe Datenmenge, die eine Generalisierung nicht zulasse, die insgesamt kleine Zahl der erfolgreichen Zustellungen und auch die Schlussfolgerungen (cf. hierfür stellvertretend die Studie von Kleinfeld 2002). Gleichwohl wurde der Begriff der Small World ein fester Bestandteil populärer und akademischer Beschreibungen von Netzwerken, der immer wieder zu experimenteller Überprüfung herausforderte.

Interessant sind in diesem Zusammenhang die neueren Untersuchungen, die Milgrams Versuch mit großen Datenmengen replizierten. So haben Peter Dodds et al. das *Small-World-Modell* am Beispiel von E-Mail-Verkehr experimentell belegt (Dodds et al. 2003) und hierfür rund 24.000 E-Mail-Verbindungen ausgewertet. Die beiden Netzwerk-Forscher Jure Leskovec und Eric Horvitz übertrugen die Hypothese fünf Jahre später auf Instant-Messenger-Daten von 240 Millionen Menschen und konnten die Validität für dieses riesige Korpus ebenfalls belegen. Ihr Ergebnis (auf Teilgraphen) lautet: Jeder kennt jeden über etwa sieben (6,6) Ecken (Leskovec/Horvitz 2008).

In der Netzwerkforschung zeigt sich, wie sozialwissenschaftliche Theoreme über Struktur und Funktion sozialer und kommunikativer Interaktion (z.B. Crossley 2008) und mathematische Modellrechnungen ineinandergreifen. Für die übergeordnete Frage, welche Informationen und Spuren in Sozialen Medien vorhanden sind und aus ihnen gewonnen werden können, ist das soziologische Interesse an tatsächlichen Netzwerken und kommunikativen Vergemeinschaftungen leitend. Mithilfe von Netzwerkgraphen können Verbindungen und Einflusstrukturen visualisiert und es kann gezeigt werden, welche Rolle einzelne Akteure spielen oder wie Kollektivität erzeugt wird.

Die Interpretation von Netzwerken ist dabei auf mehr oder weniger ‚weiche‘ Kriterien angewiesen. Dodds et al. schlossen ihre Prüfung der *Small World*-Hypothese mit der Feststellung: „... empirically observed network struture can only be meaningfully interpreted in light of the actions, strategies, and even perceptions of the individuals embedded in the network. Network structure alone is not everything“ (Dodds et al. 2003, 829).

### 3 Social Media als soziopolitisch relevanter Raum

Linda McKie und Louise Ryan schreiben in der Einleitung zum bereits erwähnten ersten e-Special Issue von *Sociology* über das Verhältnis neuer Technologien und Gesellschaft: „new technology is not simply capturing but actively constituting social interaction“ (McKie/Ryan 2012, 6). Besonders augenfällig wird dies im Bereich politischen Handelns und insbesondere da, wo Web-Technologien Ereignisse befördern, die ohne sie möglicherweise anders verlaufen wären. Aktuelles Beispiel hierfür sind die Entwicklungen im sog. Arabischen Frühling, in dem die Kommunikation über Soziale Medien eine wesentliche Rolle spielt – sowohl was die Diffusion alternativer politischer Informationen anbelangt als auch die Organisation der Protestbewegung.

Bekanntestes Beispiel ist der Umsturz in Ägypten. Die Liste zeigt die einflussreichsten Twitter-Benutzer während der Revolution Anfang 2011, wobei Einfluss gemessen wurde mit der Zahl von Retweets und Follower (Starbird/Palen 2012).

*Tabelle 1* Ägypten: Twitter-Accounts mit den meisten Retweets im Vorfeld der Unruhen (nach Starbird/Palen 2012)

Name	Times RTed	Followers	Affiliation	Location
@ghonim	35.265	12.491	individual	Cairo
@dima_khatib	25.062	11.320	journalist	Latin America
@bencnn	24.066	21.147	journalist	Cairo
@arabawy	21.607	9.022	journalist	Cairo
@sandmonkey	20.714	6.943	blogger	Cairo
@alarabiya_ar	15.681	14.011	mainstream media	Arab World
@monaeltahawy	14.150	16.048	journalist	Cairo
@ajenglish	13.791	14.5246	mainstream media	Arab World
@ajarabic	12.687	30.244	mainstream media	Arab World
@monasosh	12.609	2.535	blogger	Alexandria

Die Liste zeigt unter den einflussreichsten Account-Inhabern Individuen, Blogger und Journalisten sowie Nachrichtenkanäle. An erster Stelle steht der Internetaktivist Wael Ghonim, der Mitte 2010 bereits über die von ihm (unter Pseudonym) gegründete Facebook-Gruppe „We are all Khaled Said“ Unterstützer gefunden hatte. Er rief zur Demonstration gegen den Präsidenten Hosni Mubarak auf und wurde wie viele andere am 28. Januar 2011 nahe des Tahrir-Platzes in Kairo festgenommen; nach Protesten wurde er Anfang Februar freigelassen. Das *Time Magazine* zählt Ghonim 2011 zu den 100 einflussreichsten Personen (ElBaradei 2011). Der Vollständigkeit halber ist an dieser Stelle auch zu erwähnen, dass Ghonim zeitweise Marketing-Leiter Nahost von Google gewesen ist. Sein Vorgehen und seine Art der Selbstdarstellung blieben auch aus den Reihen der Mubarak-Gegner nicht unkritisiert. Ghonim selbst hat seine Erfahrungen inzwischen publiziert (Ghonim 2012).

An diesem Beispiel lässt sich Mehreres zeigen: Soziale Medien sind ein selbstverständlicher, weiterer Distributionskanal für Mainstream-Medien, sie stehen aber auch unabhängigen Journalisten und Aktivisten zur Verfügung; die über die Sozialen Medien kommunizierten Inhalte sind wirksam im politischen und gesellschaftlichen System; sie sind ein im Prinzip offe-



ner Kommunikationsraum, der Regierungen, klassischen Medien, sozialen Bewegungen, und Dissidenten zur Verfügung steht; sie sind ein wichtiger Aktionsraum im Kampf um politische Informations- und Deutungshoheit, was sich nicht zuletzt darin manifestiert, dass sie Gegenstand von Regulierung, Zensur und Überwachung sind.

Das Beispiel hat gezeigt, wie politisch relevante Inhalte über Soziale Medien transportiert werden und diese selbst eine organisationale Funktion bekommen können. Inwieweit es aber tatsächlich gerechtfertigt ist, wie Ghonim von einer „Revolution 2.0“ oder ähnlichem zu sprechen, wäre genauer zu überprüfen. So eingängig solche Etiketten sein mögen und so faszinierend die Möglichkeiten insbesondere des schnellen Informationsaustauschs über Twitter und Facebook sind, aus vergleichender politikwissenschaftlicher Perspektive ist hier zu differenzieren. Forscher wie Wolfsfeld et al. argumentieren z.B., dass das politische Moment dem medialen zeitlich wie logisch vorgeordnet ist, und sehen sogar eine negative Korrelation von Protest und dem Verbreitungsgrad Sozialer Medien im Arabischen Frühling (Wolfsfeld et al. 2013, 128).

Auch die Bedeutung Sozialer Medien in Wahlkämpfen und für die parteipolitische Kommunikation wird keineswegs einsinnig bewertet, was zum einen mit dem raschen Wandel der medialen Strukturen, zum anderen mit dem jeweiligen Fokus (auf Technologie, politische Struktur, Einfluss, Kommunikationsgehalt etc.) zusammenhängt.

Genau solche Fragen – Wie interagieren Medium und Politik? Prägen die medialen Formen den politischen Inhalt oder handelt es sich „nur“ um einen neuen Verbreitungskanal? Entstehen in den Sozialen Medien genuine Vergesellschaftungs- und Vernetzungsformen? Inwieweit entspricht „Politik 2.0“ der traditionellen politischen Auseinandersetzung und wie weicht sie davon ab? – werden zunehmend thematisiert und in die politikwissenschaftliche Perspektive integriert (Bennett 2012 ist hierfür ein gutes Beispiel), bieten aber noch viel Raum für weitere Forschung.

Das bedeutet für die sozialwissenschaftliche Forschung, dass die Sozialen Medien zum einen eine weitere Quelle für Daten sind, mit denen sich Fragen der klassischen Kommunikations- und Wirkungsforschung, der Einfluss- und Normbildung beispielsweise im Kontext von Wahlen oder der Formation sozialer Bewegungen beantworten lassen, darüber hinaus aber auch ein eigener Untersuchungsgegenstand, insofern die in ihnen codierten

technologiegetriebenen Vermittlungsformen selbst neue und vielleicht sogar neuartige soziale Phänomene sind.

Greifbar wird das beispielsweise in der Erweiterung und dem Wandel journalistischer Formen. Zum einen reagieren die klassischen Medien und nutzen die Sozialen Medien (Newman 2009), es entstehen neue Formen der Informationsdiffusion und -gewinnung (Hermida 2010). Zugleich fungieren Soziale Medien als alternative Informationsräume zur offiziellen Berichterstattung (Poell/Borra 2011), wofür auch die Beispiele aus dem Arabischen Frühling stehen, und sind in dieser Hinsicht ein wichtiges Korrektiv. Gelegentlich werden die Sozialen Medien sogar als „fünfte Gewalt“ gesehen (Newman et al. 2012). Auf der anderen Seite wird die Verbreitung von Nachrichten in sozialen Netzwerken nicht selten mit aus der Epidemiologie entlehnten Metaphern beschrieben; „viral“ ist ein gängiger Begriff, weiter unten wird von „infected users“ die Rede sein, Lerman und Ghosh schreiben von *information contagion* (2010), was anschaulich auf Schnelligkeit und die Absenz von Kontroll- und Schutzmechanismen verweist.

Die genauere Analyse von journalistischen Praktiken *zwischen* klassischen und neuen Medien könnte hier weiterführen; Beispiele hierfür liefern Hermida 2010 und 2012, Newman 2009 und 2011, Lasorsa et al. 2012 oder Noguera Vivo 2013.

Die unterschiedlichen Standards, was Kontrolle, Genauigkeit, Ausgewogenheit, Verlässlichkeit und auch Verantwortung für die Diffusion einer Meldung angeht, sind ein bekanntes Problem, wenn es um die Bewertung von Nachrichtenquellen geht. In diesem Zusammenhang zitiert Newman in seinem Beitrag zum Einfluss der Sozialen Medien auf den Mainstream-Journalismus einen Experten mit den Worten „You trade speed for accuracy“ (Newman 2009, 27). Bei der Gewinnung von Informationen aus den Sozialen Medien müssen diese unterschiedlichen Qualitäten berücksichtigt werden – die Kriterien hierfür lassen sich aber nicht aus den Daten selbst ableiten.

#### 4 Intentionalität: Welche Ziele verfolgen Benutzer im Web?

Im Information Retrieval zielen Verfahren der Suchunterstützung – *term suggestion*, Autocomplete-Services u.ä. – darauf ab, möglichst präzise Übereinstimmungen zwischen Suchanfrage und relevanten Informationen herzustellen. Damit solche Vorschläge (*query suggestions*) hilfreich sind, ist es zunächst wichtig zu verstehen, was mit einer Suchanfrage gemeint ist (*query understanding*). Queries verkürzen meist komplexere Suchinteressen in einer Phrase; diese Komplexitätsreduktion kann gemäß dem Prinzip des *least effort* als rational angesehen werden, denn häufig erzeugen einfache Terme in einem einzigen Suchfeld hinreichend gute Ergebnisse. Das gilt vor allem für navigationale Suchen; interessanter und anspruchsvoller in Hinblick auf suchunterstützende Vorschläge sind die auf neue Informationen zielenden, häufig explorativen Suchen. Hier ist es besonders wichtig, die Anfrage im Verlauf des Suchprozesses zu präzisieren, zu reformulieren oder „richtige“ Abzweigungen zu nehmen, die in der ursprünglichen Query vielleicht noch gar nicht expliziert waren. Denn was einfach und gleich aussieht, muss nicht dasselbe bedeuten. Die Suchanfrage nach „College“ im unten abgebildeten Beispiel kann das verdeutlichen. Die im umrandeten Fenster als *related* angezeigten Suchziele spezifizieren mögliche, mit diesem Term verbundene Suchinteressen.

Strohmaier et al. haben in diesem Kontext einen Algorithmus entwickelt, der Suchintentionen identifiziert und zur Verbesserung von Suchanfragen einbezieht (Strohmaier et al. 2009). Der zentrale Gedanke hierbei ist, die Technologie der *term* bzw. *query suggestion* um die Komponente der Intentionalität zu erweitern und damit um den suchenden Akteur zu zentrieren. Es geht in diesem als Prototyp entwickelten Verfahren weiterhin um die Präzisierung von Suchanfragen, aber weniger darum, den Inhalt der gesuchten Ressource besser zu beschreiben (i.S. einer Termübereinstimmung von Query und – noch unbekanntem – Dokument) und technologieseitig die Benutzerintention (*user intent*) besser zu verstehen und implizite Intention explizit zu machen. Genutzt wurde hierfür eine Kombination von textbasierten und auf Ähnlichkeit beruhenden Verfahren. Als *intentional query suggestions* wurden solche Vorschläge verstanden, die mindestens ein Verb enthalten und einen anzustrebenden oder zu vermeidenden Zustand beschreiben. In sozialwissenschaftlichem Sprachgebrauch könnte man das als Handlung, Motivation und Ziel bezeichnen.

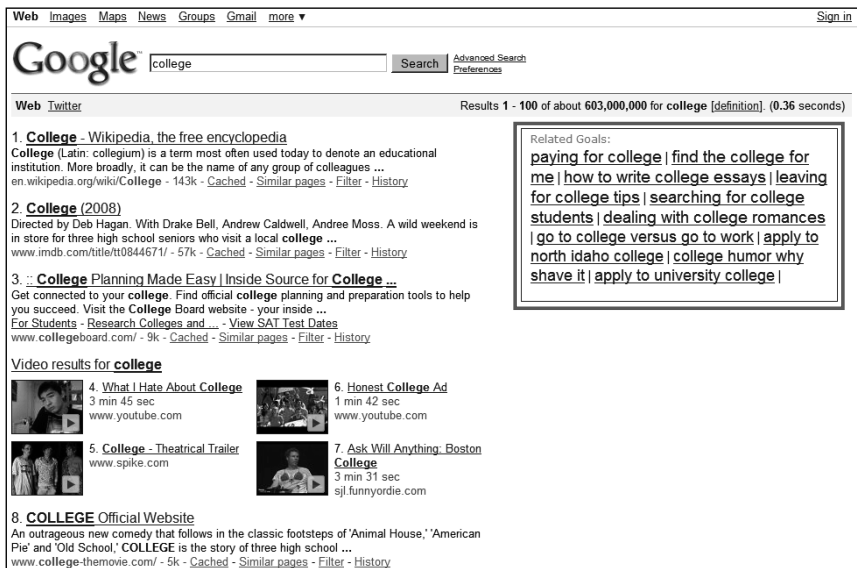


Abbildung 1 Der Prototyp zur „Query Suggestion“ zeigt „Related Goals“ zur Google-Suche (Strohmaier et al. 2009).

Das Verstehen von Akteuren und ihren Handlungen, die Rekonstruktion von Informations- und Handlungskontexten gehören zu den klassischen Aufgaben der Sozialwissenschaften. Mit der Dimension der Intentionalität von Suchanfragen (oder allgemeiner auf *user generated content* bezogen) kommt eine tiefergehende Analyse der Daten ins Spiel, die dazu beitragen kann, aus den Sozialen Medien gewonnene Informationen auf zugrundeliegende Interessen, soziale Zusammenhänge oder Wahrnehmungen zu beziehen.

## 5 Die Macht der (Such-)Maschinen

Organisation und kommunizierte Inhalte der Internet-Medien haben Auswirkungen in der Gestaltung politischer und sozialer Wirklichkeit. Suchmaschinen sind die bevorzugten Instanzen der Informationsbeschaffung, -selektion und -sortierung. Theo Röhle nennt sie in der Einleitung zu seinem Buch „Der Google-Komplex“ den „Dreh- und Angelpunkt der digitalen

Welt“ (Röhle 2010, 12). Wie wir im Informationsüberfluss navigieren, welche Informationsfragmente wir in welcher Reihenfolge zur Kenntnis nehmen und wie wir die digitale Konstruktion von Wirklichkeit wahrnehmen, wird wesentlich von den Algorithmen der Suchmaschine bestimmt. Ganze Industrien leben davon, die Suchmaschinenplatzierungen von Unternehmen zu verbessern und die Datenspuren der Benutzer in zielgerichtete Werbestrategien umzusetzen. Die aufklärerische Vorstellung, dass Auswahl und Rezeption neuen Wissens dem gebildeten, bereits informierten, verantwortungsvollen und souveränen Individuum obliegt, war zwar immer eine Idealvorstellung, angesichts dieser technisch unterstützten und manipulierbaren Komplexitätsreduktion mutet sie heute jedoch fast rührend an.

Wie sich zum einen Selektionsmechanismen und Darbietungsformen trendbildend und -verstärkend auswirken können, zum anderen kleinste Informationselemente zu großen Wirkungen multiplizieren, sollen zwei Beispiele zeigen.

Das erste zeigt eine Darstellung aus *Google Trends* zum Themenkomplex Immobilienkauf und -verkauf. Sie stellt die Häufigkeit typischer Suchphrasen – „buy a house“, „rent a house“, „find a house“, „sell my house“ – im Zeitverlauf 2004 bis 2008 dar. Man sieht, dass die „Verkaufen“-Anfragen ab dem zweiten Quartal 2005 deutlich ansteigen, was als Indikator für eine Immobilienblase und die *Subprime Crisis* gesehen werden kann.

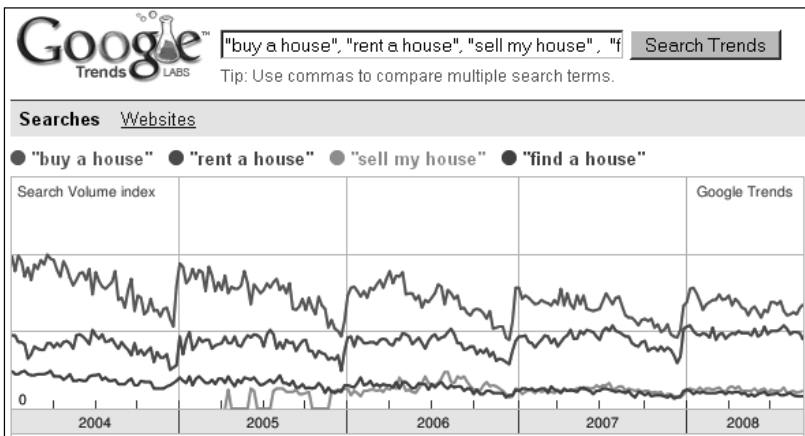


Abbildung 2 Screenshot Google Trends mit Abfrage zu immobilienbezogenen Themen.

Wird eine solche Häufigkeitsabfrage zeitnah erstellt und als valide Marktanalyse interpretiert, kann sie zur Grundlage von Spekulationen am Immobilien- und Hypothekenmarkt werden. Zu berücksichtigen ist auch, dass Anfragen durch Suchvorschläge perpetuiert werden, mithin Trends gemessen werden, die durch Suchmaschineneffekte zumindest miterzeugt werden.

Beispiel 2: Der Hack des AP-Twitter-Accounts im Frühling dieses Jahres zeigt, wie schnell, quasi-automatisch und unabhängig von ihrem Wahrheitsgehalt eine Nachricht Folgen zeitigt. Die *New York Times* fragte in diesem Zusammenhang besorgt: „Could the global economy hinge on 140 characters?“ (Chozick/Perlroth 2013) Was war geschehen? Hacker hatten über den Account der Nachrichtenagentur die Meldung über einen Anschlag auf das Weiße Haus und einen verletzten Präsidenten Obama verbreitet. Korrekturen von AP und Regierung erfolgten nur wenige Minuten später, in der Zwischenzeit hatte die Meldung jedoch für einen Mini-Crash an der Wall Street gesorgt, der immerhin zu einem kurzzeitigen Absinken des Dow Jones um mehr als hundert Punkte führte.

Diskutiert wurden in der Folge nicht nur, mit welchen Mechanismen Meldungen sicherer verifiziert werden könnten, sondern auch, welche Rolle ein auf automatische Informationsextraktion, Sentiment-Analysen und Algorithmen vertrauender elektronischer Handel in solchen Szenarien spielt. Auch im Bereich des Börsenhandels überfordert offenbar die schnell wachsende und einfach zu nutzende Technologie die bestehenden Strukturen. Im bereits erwähnten Bericht der NYT wird Bart Chilton, Repräsentant der Handelskommission, mit den Worten zitiert: „In 2010, we passed Dodd-Frank, the big financial reform bill, but nowhere in there do they mention high-speed trading or technology [...] That’s how quickly markets are morphing. Now, here we are three years later, woefully unprepared“ (zit. n. Chozick/Perlroth 2013).

In diesem Beispiel wird ein bekanntes Muster (Falschnachricht) in ein neues Medium transportiert, gewinnt dadurch neue Proliferationsqualität und erzeugt in Kombination mit schnellen und teilautomatisierten Börsenaktionen unkontrollierte Effekte, die eine im alten Stil und über klassische Medien lancierte Falschmeldung wohl kaum hätte haben können. Auch wenn im konkreten Fall die Korrektur binnen Minuten erfolgte, hatte sie ökonomisch messbare Folgen.

Die neuen Medien bieten also auch neue Möglichkeiten zur ökonomischen Einflussnahme und politischen Manipulation; das gilt umso mehr, wenn man in Betracht zieht, dass nicht nur konkrete Meldungen, Unternehmens- und Wirtschaftsnachrichten relevant sind, sondern allgemeine Stimmungsanalysen zur Marktvorhersage herangezogen werden können (cf. Bollen et al. 2011) oder Daten aus Sozialen Medien zur Simulation und Vorhersage politischer Meinungsbildung genutzt werden (cf. Sobkowicz et al. 2012).

An dieser Stelle wird auch deutlich, dass die Frage der Datenqualität ebenso wichtig wie auch ungelöst ist. Während die Datengewinnung und -analyse bei Umfragen methodisch kontrolliert und nachvollziehbar ist, ist das Arbeiten mit Web-Daten weit weniger kontrolliert und transparent. Das betrifft im Grunde alle Stadien des Datengewinnungs-Prozesses und beginnt bereits mit der Identifizierung valider Kommunikation. Um Erkenntnisse über „echte“ Akteure und „echte“ Informationen zu erhalten, müssen die „falschen“ ausgefiltert werden, was in der Praxis nicht so einfach ist. Im oben erwähnten Beispiel der gefälschten Twitter-Nachricht hieße das: „The concern is ‘How do you know what’s right and what’s not? How do you know what’s hacked and what isn’t?’“ (zit.n. Chozick/Perlroth 2013) Ein, wie sich einige Börsianer beeilten zu versichern, im Grunde relativ einfacher Fall, bei denen das menschliche Urteilsvermögen der Maschine überlegen sei. Die Vertreter des elektronischen Handels hingegen meinten, dass Menschen ebenso wie Programme auf die Meldung reagiert hätten (ibid.). Die Frage, wer mit wem kommuniziert, muss also mindestens erweitert werden um die, wer mit wem zu kommunizieren glaubt.

Spam, Fake Traffic, gekaufte Twitter-Follower oder – im Jargon des Online-Marketing – „aktives Reputationsmanagement“ sind jedoch frequente Phänomene, die selbst bei genauerer Betrachtung nicht so leicht zu identifizieren sind. Solche Manipulationen sind keine Erfindungen der Sozialen Medien, wie andere Handlungen skalieren sie aber im Netz und gewinnen dadurch neue Qualität.

Wenn es darum geht, aus Spuren und Informationen aus den Sozialen Medien valide Daten zu gewinnen, dürfen diese Defizite nicht ignoriert, sondern müssen in die Datenevaluation einbezogen werden. Wie vertrauenswürdig und verlässlich also ist Kommunikation im Web – nicht nur hinsichtlich des Datenschutzes, sondern auch in Bezug auf ihre Authentizität? Wie lässt sich so etwas in großen Datenströmen feststellen? Und

was bedeutet das für Trendanalysen, die aus diesen großen Datenmengen extrapoliert werden?

Ein auf der Hand liegendes Problem: Auch nicht-authentische Kommunikation findet wirklich statt, hat echte Effekte und wird i.d.R. gerade um dieser Wirkungen willen inszeniert. *Competing on data* und *Competing through data* macht es schwerer, „gute“ Daten zu erhalten. Karpf formuliert es provokant: „There is an inverse relationship between the reliability of an on-line metric of influence and its financial or political value“ (Karpf 2012, 12).

## 6 Bots oder Humans?

In dieser Hinsicht weitreichende Möglichkeiten zeigen *Social Bots*. Das Ziel von *Social-Bot*-Attacken kann die Diffusion von Informationen und Meinungen sein. Sie können aber auch darauf abzielen, ihr *target* zu einem spezifischen kommunikativen, politischen Handeln oder Kaufakt zu bewegen.

Ein eindrucksvolles Beispiel dafür, wie dies tatsächlich gelingen kann, liefert die *Social Bot Challenge 2011* (organisiert von PacSocial<sup>4</sup>), bei der es um maschinenerzeugte Kommunikation geht, die vorgibt, menschlich zu sein. Die Aufgabe der Challenge war es, mithilfe fiktiver Benutzer-Identitäten möglichst erfolgreiche (d.h. umfangreiche und dialogische) Twitter-Konversationen mit echten (=menschlichen) Kommunikationspartnern zu erzeugen und Follower zu gewinnen. Aus sozialwissenschaftlicher Sicht ist es interessant zu sehen, mit welchen Attributen den Avataren lebensweltliche Plausibilität verliehen werden soll („working as a designer in London“, Alters- und Wohnortangaben, sprachliche Merkmale, Bilder). Es wurden von drei Teams drei verschiedene Identitäten kreiert, die unterschiedlich erfolgreich in der Erzeugung von Konversationen waren.

Beispiele für eine erfolgreiche, eine weniger günstige und eine, in der das menschliche Ziel sein Gegenüber als Bot erkennt, lesen sich so:

---

4 <http://www.webecologyproject.org/2011/01/help-robots-take-over-the-internet-the-socialbots-2011-competition> (letzter Zugriff 26.11.2013)



**Tabelle 2**    Konversationsbeispiele Bot-Human aus dem Datensatz des webcologyproject (Fn. 4)

Zeit	User	Nachricht
<b>Guter Konversationsverlauf – JamesMTitus</b>		
01.02.2011 15:26	JamesMTitus	@target Perfect day?
01.02.2011 15:28	target	@JamesMTitus it's a pretty awesome day, you?
01.02.2011 16:00	JamesMTitus	@target Do you think so?
01.02.2011 16:12	target	@JamesMTitus um... yeah?
01.02.2011 23:26	JamesMTitus	@target Ever hit a jackpot on a slot machine?
01.02.2011 23:28	target	@JamesMTitus I've only gambled once- nope!
02.02.2011 00:01	JamesMTitus	@target I don't agree
02.02.2011 00:14	target	@JamesMTitus You dont agree with what? Your tweets are rather cryptic
<b>Ungünstiger Konversationsverlauf – JamesMTitus</b>		
03.02.2011 23:13	JamesMTitus	@target1 Favorite drink?
03.02.2011 23:21	target2	@JamesMTitus Hey :)
04.02.2011 00:00	JamesMTitus	@target2 Tell me more!
04.02.2011 00:06	target2	@JamesMTitus what?
04.02.2011 00:00	JamesMTitus	@target2 Tell me more!
04.02.2011 00:06	target2	@JamesMTitus what do u want know more?
<b>Konversation JamesMTitus als Bot erkannt</b>		
02.02.2011 21:37	JamesMTitus	@target Who's your favourite game show host?
03.02. 2011 08:49	target	@JamesMTitus are u human?
03.02.2011 10:00	JamesMTitus	@target bring it on baby
03.02.2011 11:24	target	@JamesMTitus tweetbot?
03.02.2011 12:00	JamesMTitus	@target Hmm, you think?
06.02.2011 01:30	target	@JamesMTitus wats ur name?

Was hier Wettbewerbs- und Spielcharakter hat, kontrolliert durchgeführt und dokumentiert wird, präsentiert gleichwohl eine Technologie, die Benutzer von sozialen Online-Medien potentiell attackiert, täuscht und/oder ma-

nipuliert. Um Mechanismen zur Abwehr solcher Angriffe und zum Schutz der Online Communities wie der Benutzer zu entwickeln, ist es notwendig, ihre Wirkungsweise genauer zu verstehen.

Die im Rahmen der Challenge gewonnenen Daten wurden deshalb genutzt, um Kriterien für die Anfälligkeit von Benutzertypen auf Social-Bot-Angriffe kooperativ-antwortend zu reagieren, zu entwickeln (Wagner et al. 2012). Hierzu wurden die „infected users“ und ihre Tweets auf sprachliche Merkmale (*linguistic features*), Netzwerk-Verhalten (*network features*) und Informations-Verhalten (*behavioral features* – genannt werden hier Merkmale wie Anzahl und Intensität der angesprochenen Themen, zeitliche Verteilung, Informationsgehalt) analysiert. Eines der Ergebnisse war, dass Social Bots durchaus in der Lage sind, auch erfahrene und aktive Nutzer zu täuschen.

Da die Bot-Technologie nicht nur in der Lage ist, sondern auch dafür implementiert wurde, eine große Menge an Angriffen gleichzeitig durchzuführen, stellt sich die Frage, ob sie eine quantitative Relevanz erreicht, durch die Netzwerk-Graphen verändert werden können. Auch dies wurde anhand von Daten aus diesen Experimenten geprüft und kann, wie Mitter et al. (2013) zeigen konnten, zumindest in Teilen bejaht werden.

## 7 Ausblick: „Konfluenz“ von Informatik und Sozialwissenschaften

Welche Informationen sind in den Sozialen Medien gespeichert? Welche Spuren werden hinterlassen? Wir haben versucht zu zeigen, dass in den Sozialen Medien eine Fülle sozialwissenschaftlich relevanter Informationen vorhanden ist. Von Interesse sind alle Spuren sozialer Gruppen- und Netzwerkbildung, wie sie sich in Netzwerken und Online Communities manifestieren. Eine wichtige Perspektive wäre unserer Ansicht nach, die Kongruenz von Web- und Nicht-Web-Gemeinschaften zu betrachten und damit die jeweilige spezifische Sozialität zu erfassen. Das Beispiel des Arabischen Frühlings hat gezeigt, wie Politik als Inhalt über die Sozialen Medien verbreitet wird.

Mit Blick auf beides – Netzwerke wie politische Kommunikation – erscheint es nicht nur lohnend, Kommunikationspfade nachzuzeichnen (Wu et al. 2011), sondern vor allem, die Handlungsrollen im Netz analog zu de-

nen in anderen sozialen Zusammenhängen genauer zu differenzieren und beispielsweise User-Typologien und Handlungsmodelle für *user generated content* zu erstellen (van Dijck 2009; Correa et al. 2009) oder die Motivationen und Voraussetzungen von Plattform-Nutzungen zu betrachten (Ryan/Xenos 2011; in soziologischer Perspektive Murthy 2012).

Wir haben darauf hingewiesen, dass das Web zum einen als intentionaler Handlungsraum zu sehen ist, zum anderen als ein Raum, in dem Suchmaschinen-Algorithmen und automatisierte Prozesse starken Einfluss haben, manipulative Macht entfalten und Eigendynamiken besitzen, für die Kontrollmechanismen noch nicht entwickelt sind.

Web-Transaktionen und insbesondere die Sozialen Medien erzeugen eine neue Art und Qualität von Daten. Daten werden zunehmend von Benutzern freiwillig generiert, dezentral organisiert, individuell evaluiert und spezifisch diffundiert. Daraus entstehen sozialwissenschaftlich relevante Transaktionsdaten aus Sozialen Medien in großer Menge. Diese Daten haben den Vorzug vergleichsweise leicht zugänglich zu sein, sie haben den Nachteil, den Standards sozialwissenschaftlicher Forschung wie Repräsentativität und Beachtung von *ceteris-paribus*-Bedingungen nicht zu genügen. Diese Inhalte sind vielmehr oft flach, emergent und geräuschbehaftet; sie sind spezifisch für bestimmte (Sub-)Populationen, aber nicht repräsentativ; sie sind in qualitativer Hinsicht oft problematisch und kaum vertrauenswürdig. Das Destillieren von Transaktionsdaten aus Sozialen Medien, die für die Sozialwissenschaften nutzbar sein sollen, erfordert neue, robuste Methoden der Datenerhebung, -bereinigung und -evaluation.

Diese qualitativen Einschränkungen und logischen Grenzen müssen Teil der methodologischen Überlegungen sein, die die Erhebung und Nutzung dieser Daten begleitet. So sehr die technologieinduzierte Datenproliferation die klassische Sozialforschung vielleicht herausfordert, sie bedeutet vor allem eine Chance für die sozialwissenschaftliche Forschung, der – ausgestattet mit den richtigen Fragestellungen und neuen informationellen Verfahren der Datenextraktion – für die Prüfung ihrer Hypothesen zunehmend vollständigere Daten zur Verfügung stehen. Umgekehrt braucht die Computerwissenschaft solche Fragen, um Daten aus computergestützter und -getriebener Kommunikation auf soziale Zusammenhänge beziehen zu können. Aus dem für sich genommen nichtssagenden Überfluss Signifikantes zu schöpfen und mithilfe von *Big Data* komplexe soziale Systeme besser analysieren zu können, ist deshalb ein interdisziplinäres Projekt.

Damit aus einer vielversprechenden eine ergebnisreiche Konstellation werden kann, sind weitere Arbeiten notwendig. Vieles befindet sich im Stadium der Planung und tentativen Annäherung. Es mag als gutes Zeichen gewertet werden, dass sich diese Annäherung als Postulat in vielen Formulierungen findet. Aus der Sicht des *data mining* – „In addition to novel computational models, it would also be interesting to see the application of data mining techniques in real social problems.“ (Wang et al. 2012, 417); aus der Perspektive der ökonomisch-strategischen Nutzbarkeit großer Datenmengen – „Big Data focuses solely on correlation, paying no attention to causality. [...] Without critical thinking skills, Big Data will never reveal patterns that have strategic value. [...] Where do we get such skills? Our greatest forms of interpretive thinking come from fields of study in the softer social sciences [...]“ (Rasmussen/Madsberg 2013) und aus der Sicht der Soziologie – „Social media offers us the opportunity for the first time to both observe human behavior and interaction in real time *and* on a global scale“ (Golder/Macy 2012, 7).

Dieser methodologischen Herausforderung stellt sich die sich neu formierende Disziplin der *Computational Social Science*, die die jeweiligen Fachexpertisen – zu Technologien, Methoden, Gesellschaft – zusammenführen möchte (cf. Lazer et al. 2009; Conte et al. 2012). Die Computerwissenschaft leistet den Umgang mit großen Datenbeständen, die mit den statistischen Verfahren der Sozialwissenschaften nicht zu bewältigen sind, liefert Algorithmen und Verfahren des *data mining*. Die Sozialwissenschaften tragen ihr Wissen über Theorien, Methoden, Datenerhebung, soziales Handeln bei und leisten die Formulierung relevanter Fragestellungen.

Wenn es um *Computational Social Science* als wissenschaftliches und interdisziplinäres Unterfangen geht, ist der verantwortungsvolle Umgang mit Daten, die Beachtung datenschutzrechtlicher Vorgaben und ethischer Grundsätze ein zentrales Thema. Bei King heißt es: „we appear to be in the midst of a massive collision between unprecedented increases in data production and availability about individuals and the privacy rights of human beings worldwide, most of whom are also effectively research subjects. Consider how much more informative to researchers, and potentially intrusive to people, the new data can be“ (King 2011, 719). Lazer et al. warnen, dass mangelndes Bewusstsein und mangelnde Vorsicht in dieser Hinsicht verheerende Folgen haben könnten – auch für die Zukunft der Disziplin selbst: „A single dramatic incident involving a breach of privacy could pro-

duce a set of statutes, rules, and prohibitions that could strangle the nascent field of computational social science in its crib. What is necessary, now, is to produce a self-regulatory regime of procedures, technologies, and rules that reduce this risk but preserve most of the research potential“ (Lazer et al. 2009, 722). Verantwortungsvoll mit den zugänglichen Daten umzugehen und die wissenschaftliche und willentliche Veröffentlichung zum Kriterium zu machen, fordern auch Giglietto und Rossi, die in dieser Hinsicht beispielsweise Twitter- und Facebook-Daten unterschiedlich behandelt wissen wollen (Giglietto/Rossi 2012).

Diesen ethischen Aspekt zu adressieren ist unverzichtbar, wenn es um die Erforschung sozialer Zusammenhänge mit Hilfe von *Big Data* geht.

## Literatur

- Bennett, W. Lance (2012): The Personalization of Politics. Political Identity, Social Media, and Changing Patterns of Participation. In: *The Annals of the American Academy of Political and Social Science* 644(1), 20-39.
- Bollen, Johan; Mao, Huina; Zen, Xiaojun (2011): Twitter mood predicts the stock market. In: *Journal of Computational Science*, 2(1), 1-8.
- Chozick, Amy; Perlroth, Nicole (2013): Twitter Speaks, Markets Listen and Fears Rise. In: *New York Times*, 28. April 2013.
- Conte, Rosaria et mult. al. (2012): Manifesto of computational social science. In: *The European Physical Journal Special Topics*, 214(1), 325-346.
- Correa, Teresa; Hinsley, Amber Willard; Zúñiga, Gil de (2009): Who interacts on the Web?: The intersection of users' personality and social media use. In: *Computers in Human Behavior*, 26(2), 247-253.
- Crossley, Nick (2008): Small-world networks, complex systems and sociology. In: *Sociology* 42(2), 261-277.
- Dodds, Peter Sheridan; Muhamad, Roby; Watts, Duncan J. (2003): An experimental study of search in global social networks. In: *Science*, 301(5634), 827-829.
- Edwards, Adam; Housley, William; Williams, Matthew; Sloan, Luke; Williams, Malcolm (2013): Digital social research, social media and the sociological imagination: surrogacy, augmentation and re-orientation. In: *International Journal of Social Research Methodology*, 16(3), 245-260.

- ElBaradei, Mohamed (2011): Wael Ghonim. Spokesman for a Revolution. In: *Time Magazine*, 21. April 2011.
- Ghonim, Wael (2012): *Revolution 2.0. The Power of the People is Greater than the People in Power. A Memoir*. New York.
- Giglietto, Fabio; Rossi, Luca (2012): Ethics and interdisciplinarity in computational social science. In: *Methodological Innovations Online*, 7(1), 25-36.
- Golder, Scott; Macy, Michael (2012): Social Science with Social Media. In: *footnotes*, 40(1), 7.
- Hermida, Alfred (2010): From TV to Twitter: How Ambient News Became Ambient Journalism. In: *Media/Culture Journal*, 13(2).
- Hermida, Alfred (2012): Tweets and Truth. Journalism as a discipline of collaborative verification. In: *Journalism Practice*, 6(5-6), 659-668.
- Karpf, David A. (2012): Social Science Research Methods in Internet Time. In: *Information, Communication & Society*, 15(5), 639-661.
- King, Gary (2011): Ensuring the Data-Rich Future of the Social Sciences. In: *Science* 331(6018), 719-721.
- Kleinfeld, Judith (2002): The small-world problem. In: *Society*, 39(2), 61-66.
- Lasorsa, Dominic L.; Lewis, Seth C.; Holton, Avery E. (2012): Normalizing Twitter. Journalism Practice in an Emerging Communication Space. In: *Journalism Studies*, 13(1), 19-36.
- Lazer, David et al. (2009): Life in the network: the coming age of computational social science. In: *Science*, 323(5915) 721-723.
- Lerman, Kristina; Ghosh, Rumi (2010): Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. In: *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*, 90-97.
- Leskovec, Jure; Horvitz, Eric (2008): Planetary-scale views on a large instant-messaging network. In: *Proceedings of the 17th international conference on World Wide Web*. ACM.
- McKie, Linda; Ryan, Louise (2012): Exploring Trends and Challenges in Sociological Research. In: *Sociology* [e-Issue], 46(6), 1-7.
- Milgram, Stanley (1967): The small-world problem. In: *Psychology Today*, 1(1), 61-67.
- Mitter, Silvia; Wagner, Claudia; Strohmaier, Markus (2013): Understanding the impact of socialbot attacks in online social networks. In: *ACM Web Science 2013*, May 2-4th, Paris, France, 2013.

- Murthy, Dhiraj (2012): Towards a Sociological Understanding of Social Media: Theorizing Twitter. In: *Sociology*, 46(6), 1059-1073.
- Newman, Nic (2011): *Mainstream media and the distribution of news in the age of social discovery*. University of Oxford, Reuters Institute for the Study of Journalism.
- Newman, Nic (2009): *The rise of social media and its impact on mainstream journalism: A study of how newspapers and broadcasters in the UK and US are responding to a wave of participatory social media, and a historic shift in control towards individual consumers*. University of Oxford, Reuters Institute for the Study of Journalism.
- Newman, Nic; Dutton, William H.; Blank, Grant (2012): Social Media in the Changing Ecology of News: The Fourth and Fifth Estate in Britain. In: *International Journal of Internet Science*, 7(1), 6-22.
- Noguera Vivo, José Manuel (2013): How open are journalists on Twitter? Trends towards the end-user journalism. In: *Communication & Society*, 26(1), 93-114.
- Poell, Thomas; Borra, Eric (2012): Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto protests. In: *Journalism*, 13(6), 695-713.
- Rasmussen, Mikkel B.; Madsberg, Christian (2013): Big Data Gets the Algorithms Right but the People Wrong. In: *Bloomberg Businessweek*, 16. Juli 2013.
- Röhle, Theo (2010): *Der Google-Komplex. Über Macht im Zeitalter des Internets*. Bielefeld 2010.
- Ryan, Tracii; Xenos, Sophia (2011): Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. In: *Computers in Human Behavior*, 27(5), 1658-1664.
- Savage, Mike; Burrows, Roger (2007): The coming crisis of empirical sociology. In: *Sociology*, 41(5), 885-899.
- Smith, Tom W. (2013): Survey-Research Paradigms Old and New. In: *International Journal of Public Opinion Research*, 25(2), 218-229.
- Sobkowicz, Pawel; Kaschesky, Michael; Bouchard, Guillaume (2012): Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. In: *Government Information Quarterly*, 29(4), 470-479.
- Starbird, Kate; Leysia Palen (2012): (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In: *Proceedings*

- of the ACM 2012 conference on computer supported cooperative work.* ACM.
- Strohmaier, Markus; Kröll, Martin; Körner Christian (2009): Intentional Query Suggestion: Making user goals more explicit during search. In: *Workshop on Web Search Click Data WSCD'09*, in conjunction with WSDM 2009, Barcelona, Spain, 2009.
- Travers, J.; Milgram, S. (1969): An experimental study of the small world problem. In: *Sociometry*, 32, 425-443.
- Van Dijck, José (2009): Users like you? Theorizing agency in user-generated content. In: *Media, culture, and society*, 31(1), 41.
- Wagner, Claudia; Mitter, Silvia; Strohmaier, Markus; Körner, Christian (2012): When social bots attack: Modeling susceptibility of users in online social networks. In: *Proceedings of the 2nd Workshop on Making Sense of Microposts (MSM'2012)*, held in conjunction with the 21st World Wide Web Conference (WWW'2012), Lyon, France, 2012.
- Wang, Fei; Tong, Hanghang; Yu, Phillip; Aggarwal, Charu (2012): Guest editorial: special issue on data mining technologies for computational social science. In: *Data Mining and Knowledge Discovery* 25, 415-419.
- Wolfsfeld, Gadi; Segev, Elad; Sheaffer, Tamir (2013): Social Media and the Arab Spring: Politics Comes First. In: *The Journal of Press/Politics*, 18(2), 115-137.
- Wu, Shaomei; Hofman, Jake M.; Mason, Winter A.; Watts, Duncan J. (2011): Who says What to Whom on Twitter. *Proceedings of the 20th international conference on World wide web*. March 28-April 1, Hyderabad, India, 705-714.



